

Faktantarkistuksen jälkeinen aika: joukkoälyä vai joukkohulluutta?

Nykyinen sosiaalinen media vaikuttaa ilmentävän joukkoälyn pimeitä puolia, ja poliittinen polarisoituminen, vihamieliset toimijat, sosiaalisen median taloudelliset yllykkeet sekä tekoälyn käyttö uhkaavat sitä vähää joukkoälyä mitä sosiaalisessa mediassa on jäljellä. Tarjoaisiko yhteisöhuomautusjärjestelmä joukkoistettuna faktantarkastuksena toimivan vaihtoehdon faktantarkistuksen jälkeiseen aikaan?

TEKSTI: OSKARI LAPPALAINEN

Tarkastelen artikkelissani faktantarkistusta ja sen saamaa kritiikkiä, yhteisöhuomautusjärjestelmää joukkoistettuna faktantarkistuksena ja sen suhdetta joukkoälyyn sekä sosiaalisen median tulevaisuutta laajemmin joukkoälykeskustelun valossa.

Tammikuussa 2025 sosiaalisen median jättiyhtiö Meta yllätti ilmoittamalla, että se lopettaa kolmansia osapuolia käyttäneen faktantarkistusohjelmansa. Ohjelma sai alkunsa vuonna 2016, kun Facebook joutui rankan kritiikin kohteeksi siellä levinneen disinformatio-ongelman vuoksi. Nyt yhtiö tekee täyskäännöksen, ja osa ehdotetuista muutoksista on eurooppalaisesta perspektiivistä käytännössä vihapuheen sallimista.

Metan liike on mitä selkeimmin kostonhimoa puhkuvan Donald Trumpin uudelleenvalinnan aiheuttama, ja Zuckerberg on historiallisesti liikkunut sinne, minne poliittinen tuuli on puhaltanut. Zuckerbergin tekemä Metan uudet linjaukset sisältävä julistus kuitenkin heijastaa laajempaa Yhdysvaltalaisen kokemaa epäluottamusta ja tyytymättömyyttä faktantarkistusta kohtaan. Kolmansien osapuolien tuottaman

faktantarkistuksen korvaa nykyään X:ssä käytössä oleva yhteisöhuomautusjärjestelmä.

Faktantarkistuksessa on monia haasteita

Faktantarkistusta voidaan tehdä joko ennen julkaisua tai julkaisun jälkeen. Erilaisten faktantarkistusorganisaatioiden julkaisun jälkeen suorittamat tarkistukset nousivat yleiseen tietoisuuteen erityisesti vuoden 2016 Yhdysvaltojen presidentinvaalien jälkeen, jonka aikana perattiin erityisesti Trumpin kampanjan leviämää virheellistä tietoa.

Nykymuotoista faktantarkistusta on kuitenkin kritisoitu koko sen olemassaolon ajan. Suurin kritiikki on liittynyt siihen, että se on koettu poliittisesti motivoituneeksi yritykseksi puolustaa konsensusta tai organisaatioiden jäsenten mielipiteiden mukaista politiikka. Faktantarkistus onkin Yhdysvalloissa koettu muuttuneen enemmän mielipidejournalismiksi.

Faktantarkistuksen on myös koettu olevan elitististä ja vetoavan auktoriteettiargumentaatioon, jossa tarkastusta tekevä lähde on esitetty lähtökohtaisesti vastaanottavaa yleisöä asiasta paremmin tietäväksi. Tämän on havaittu aiheuttavan vasta-reaktiota, ja jopa olemassa olevien asenteiden koven tumista. Faktantarkistuksen kyky vaikuttaa ihmisiin, jotka ovat jo päättäneet mielipiteensä vaikuttaakin olevan varsin rajallinen.

Faktantarkistuksen haasteena on internetissä syntyvän sisällön volyymi. Faktantarkistajat eivät ehdi reagoida kaikkeen, ja yritykset tehdä näin ovat tuottaneet sävelkuuroja ”faktantarkistuksia” poliittiselle satiirille tai keskusteluille, joihin oikean ja väärän tiedon käsitteet eivät istu.



**Faktantarkistus
on Yhdysvalloissa
koettu muuttuneen
enemmän mielipide-
journalismiksi.**

Faktantarkistus ei siis ole ongelmaton. Voisiko Metan korvaavaksi järjestelmäksi kaavailemat yhteisöhuomautukset olla tällainen järjestelmä?

Toimiiko moderointi yleisöhuomautuksilla?

Yhteisöhuomautukset (Community Notes) ovat X-alustan yhteisölähtöinen moderointityökalu, jossa käyttäjät voivat lisätä kontekstia julkaisuihin, kuviin ja videoihin. Ominaisuus pyrkii torjumaan harhaanjohtavaa sisältöä käyttämällä algoritmia, joka priorisoi monipuolisista näkökulmista hyödylliseksi arvioituja. Se sai alkunsa vuonna 2020 nimellä Birdwatch, eli kyseessä ei ole Muskin ajan innovaatio. Yhteisöhuomautukset ovat toisin sanoen oikeastaan joukkoistettua faktantarkistusta.

Kun olin vielä X:ssä, näin usein yhteisön huomautuksia mainosten yhteydessä. Myös tekoälyllä generoidut kuvasisällöt ja selkeät huijaukset, jotka eivät liittyneet päivänpoliittisiin kuumiin perunoihin saivat usein huomautuksia. Kaikesta polarisaatiosta huolimatta ihmiset tuntuivat löytävän jonkinlaista konsensusta tietyistä teemoista.

Tutkimusten mukaan ohjelma on parantanut sisällön tarkkuutta, mutta toisaalta se on kaukana ongelmattomasta. Keskeinen haaste liittyy X:n viestisällön volyymiin. Esimerkiksi Center for Countering Digital Hate:n tutkimusten mukaan 74 % Yhdysvaltain vuoden 2024 vaaleihin liittyvästä väärästä tiedosta jäi ilman huomautusta.

Toinen ongelma liittyy siihen, että yhteisön huomautusten vaatiman ”konsensuksen” laaja muodostuminen vie aikaa, jos siihen edes päästään. Ainoastaan 10% ehdotetuista huomautuksista tulee julkaisuksi. Atlantic Councilin tutkimuksen mukaan järjestelmä toimii hitaasti ja epäjohtamukaisesti nopean uutisyklin tapahtumissa. Analyysi 400 harhaanjohtavasta julkaisusta osoitti, että huomautuksen ilmestymiseen kului keskimäärin 7 tuntia, joissakin tapauksissa jopa 70 tuntia. Sekä perinteinen faktantarkistus että yhteisöhuomautukset siis kohtaavat aivan samoja informaation volyymiin ja syklien nopeuteen liittyviä ongelmia.

Joukkoälyä vai yhdenmukaistavaa laumautumista?

Joukkoälyllä viitataan ryhmässä olevien itsenäisten yksilöiden myötä syntyneeseen yhteiseen tietämykseen. James Surowieckin kirjaa *The Wisdom of the Crowds* vuodelta 2004 voidaan pitää käsitteen modernina artikulaatioina, mutta se ei ole ideana uusi. Surowiecki avaa kirjansa Francis Galtonin anekdootilla vuodelta 1907, jolloin markkinoille kokoontuneen ihmisjoukon antamien arvioiden keskiarvo osoittautui esillä olleen teurastetun härän oikeaksi painoksi.

**Faktantarkistuksen
kyky vaikuttaa
ihmisiin on
varsin rajallinen.**



Surowieckin pääteesi on se, että ryhmien tekemät päätökset ovat usein parempia kuin yksittäisten jäsenten. Joukkoälyn edut

Surowiecki jakaa kolmeen päätyyppiin: kognitioon, koordinointiin ja yhteistyöhön. Surowiecki näkee, että joukkoistettu ajattelu ja tiedonkäsittely voivat olla nopeampi, luotettavampia ja vähemmän alttiita poliittisille vaikutteille kuin esimerkiksi komiteoiden päätökset.

Ihmisjoukot kykenevät myös tekemään huonoja kollektiivisia päätöksiä. Näitä ovat esimerkiksi osake-markkinoiden kuplat. Surowiecki näkee joukkoälyn huonojen päätösten syntyvän liiasta konformismista ja emotionaalisuudesta. Optimaaliseen joukkoälyyn kykenevä joukko sallisi mielipiteiden monipuolisuuden, ihmisten itsenäisyyden, luottamuksen ja hajauttamisen, mutta toisaalta myös mekanismin, jolla koottaisiin yksityiset arviot yhdeksi päätökseksi.

Kirjan julkaisuvuosi 2004 oli tietoteknologian optimismin kulta-aikaa, ja on helppoa nähdä miksi uusien tieto- ja viestintäteknologioiden tuomat mahdollisuudet loihdivat esiin optimistisia näkemyksiä joukkoälystä. Wikit, blogit ja varhainen Google olivat luoneet positiivisia esimerkkejä joukkoälystä. Surowiecki vaikuttaa kuitenkin ounastelleen myös kielteisiä kehityskulkuja. Vuonna 2005 O'Reilly Emerging Technology -konferenssissa hän pohti, onko mahdollista, että ihmisryhmistä tulee teknologian kehityksen myötä liian yhteenkiinnittyneitä. Tällöin riskiksi nousevat informaatioteorian ennustamat informaatiokaskadit, jossa ihmiset päätyvät tekemään samanlaisia päätöksiä yksi toisensa jälkeen laumanomaisesti.

>>

**Sekä perinteinen
faktantarkistus että
yhteisöhuomautukset
kohtaavat samoja
informaation volyymiin
ja syklien nopeuteen
liittyviä ongelmia.**



**Kukaan ei uhkaa
aseella pysymään
X:ssä tai Metan
palveluissa.**

Sosiaalisen median moderoinnin rajat

Miltä maailma näyttää nyt 11 vuotta *The Wisdom of the Crowdsin* julkaisun jälkeen? Voidaan varmuudella sanoa, että emme ole päätyneet minkäänlaiseen suowiekiläiseen utopiaan. Enemmän vaikuttaa siltä, että hänen kuvaamat negatiiviset uhkakuvat ovat toteutuneet. Sosiaalinen media on tuottanut polarisaatiot, missä eri tavoin ajattelevat ovat vahvasti leiriytyneet keskenään. Puolensa valinneiden keskuudessa ei juurikaan suvaita yrityksiä ymmärtää ”niiden muiden” ajattelua tai argumentointia.

Se, että yhteisöhuomautusten kaltainen järjestelmä kuitenkin toimii joissain tapauksissa, antaa toivoa siitä, että emme ole myöskään täysin vajonneet jonkinlaisen totaalisen joukkotyhmyyden dystopiaankaan. Älykäs joukko saattaa yksinkertaisesti olla enemmän poikkeus kuin normi, ja toisaalta viimeisen 11 vuoden aikana on tehty hyvin vähän mitään, mikä saattaisi edesauttaa tällaisten joukkojen syntymistä. Yhteisöhuomautukset saattavat siis olla terve askel tähän suuntaan.

Kaikkein fundamentaalisin ongelma joukkoälyssä ja yhteisöhuomautuksissa on yksinkertaisesti se, että se nojaa ideaan jaetusta, yhteisestä ”joukosta” tai ”yhteisöstä”, jossa kaikki toimijat toimivat hyvässä uskossa – tai ovat edes ihmisiä. Poliittisesti ja kaupallisesti motivoituneilla tahoilla on voimakkaat intressit pyrkiä muokkaamaan narratiiveja, ja jopa maksaa ihmisille tästä. Nykyaikainen tekoälyteknologia on mahdollistamassa sen, että enää ei edes tarvita ihmisiä tekemään vaikutustyötä.

Loppujen lopuksi päätös siirtyä kohti yhteisöhuomautuksen kaltaisia järjestelmiä on poliittisten ja taloudellisten intressien yhteentulema. Elon Musk on teknologiaoligarkeista kaikkein selvimmin

poliittisesti motivoitunein, mutta vaikuttaa siltä, että puhe inklusiosta on ollut hänen vertaisilleen lähinnä ikävä velvoite jolla on myötäilty senhetkistä poliittista konsensusta. Lisäksi on muistettava, että sisällön moderointi maksaa, olkoonkin että teknologiayritykset ovat pyrkineet minimoimaan siihen käytetyt kustannukset. Esimerkiksi Facebook on ulkoistanut moderointiaan Kenian kaltaisiin maihin, jossa ihmiset ovat joutuneet käsittelemään syvästi traumatisoivaa sisältöä 1.50 dollarin tuntipalkalla.

Oletettavasti seuraava peliliike on jonkinlainen tekoälyllä toteutettu sisällönhallinta. Periaatteessa tällä voitaisiin saavuttaa merkittäviä erävoittoja esimerkiksi ihmisten altistumista traumatisoivaa sisältöä vastaan. Toisaalta tällaiset järjestelmät tarkoittavat yksinkertaisesti sitä, että se joka pääsee määrittelemään niiden toiminnan, pääsee määrittelemään ihmisten koetun konsensustodellisuuden.

Yhtiövetoisen sosiaalisen median maailmassa tämä tarkoittaa sitä, kuka maksaa eniten.

Järjestelmä ei ole enää korjattavissa

Moderoinnin mahdottomuus heijastaa tiettyssä mielessä laajemmin sosiaalisen median mahdottomuutta. Alan yrityksillä on ollut suuria vaikeuksia muuttaa toimintaansa tuottavaksi liiketoiminnaksi. Mainos-tulot ovat olleet keskiössä. Poliittiset vaikuttajat ovat myös mainostajia, joilla on suuria intressejä vaikuttaa ihmisten käyttäytymiseen. Poliitiikka on tunnetusti mahdollisuuksien, ei totuuksien, taidetta, ja nykyisillä poliittisilla järjestelmillä ei välttämättä ole intressejä kultivoida sellaisia valistuneita kansalaisia, jotka kykenisivät muodostamaan älykkäitä joukkoja.

Koska sosiaaliset mediat ovat käytännössä kasvaneet riippuvuusmekanismeilla toimiviksi mielipidevaikutusalustoiksi, ei ihmisillä ole suuria intressejä lähteä maksamaan niiden käytöstä. Tämä tekee mahdottomaksi järjestelmien reformin vähemmän manipulatiivisiksi.

Kaiken kaikkiaan keskustelut sosiaalisen median moderoinnista, faktantarkastuksesta ja yhteisöjen huomautuksista sivuuttavat tämän nykyisten järjestelmien lähtökohtaisen mahdottomuuden, ja sen, että nämä järjestelmät eivät ole enää korjattavissa.

Parasta mihin voimme siis tähdätä, on sosiaalisen median jälkeinen aika. Vaikka Yhdysvallat Trumpin johdolla harjoittaisi millaista politiikkaa tahansa, jokainen voi omana yksilönään poistua sosiaalisen median järjestelmistä. Vielä kukaan ei uhkaa aseella pysymään X:ssä tai Metan palveluissa. ■

Kirjoittaja työskentelee asiantuntijana Laurea AMK:ssa. Hän on kiinnostunut ihmisten ja teknologian kanssakäymisestä sekä kestävyyydestä ja resilienssistä laajasti ymmärrettynä.

**Parasta
mihin voimme
siis tähdätä, on
sosiaalisen median
jälkeinen aika.**